

Bias

Ashwini Kalantri

Moderator: Dr BS Garg

“Any systematic error in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure’s effect on the risk of the disease.”

–Leon Gordis

Types

- Selection bias
- Information bias/Measurement bias
- Bias due to confounding

Selection Bias

- Selection bias is a systematic error resulting from the way the subjects are either selected in a study or else are selectively lost to follow up.
- Selection bias can cause an overestimate or underestimate of association.

Selection Bias example

Study of asbestos exposure and lung cancer, the exposure is distributed among cases and controls in target population

	Diseased	Non diseased
Exposed	100	200
Unexposed	100	400

True OR in target population = $(100 \times 400) / (100 \times 200)$
= 2

Selection Bias

example

If the selection probabilities for all the cells in the table are equal at 90%, the 2x2 table of selection probabilities would be:

	Diseased	Non diseased
Exposed	$100 \times 0.90 = 90$	$200 \times 0.90 = 180$
Unexposed	$100 \times 0.90 = 90$	$400 \times 0.90 = 360$

$$\begin{aligned} \text{OR} &= (90 \times 360) / (90 \times 180) \\ &= 2 \end{aligned}$$

Selection Bias **example**

If selection probabilities are unequal, and non proportional, then selection bias will occur

	Diseased	Non diseased
Exposed	$100 \times 0.90 = 90$	$200 \times 0.90 = 180$
Unexposed	$100 \times 0.90 = 90$	$400 \times \mathbf{0.70} = 280$

$$\begin{aligned}\text{OR} &= (90 \times 280) / (90 \times 180) \\ &= \mathbf{1.6}\end{aligned}$$

Self-selection Bias

- Common source of selection bias
- Volunteers induced bias
- Individuals who volunteer for study possess different characteristics than average general population
- **Example:** A case control study explored an association of family history of heart disease and presence of heart disease in subjects. Volunteers were recruited. Subjects with heart disease may be more likely to participate if they have family history

Self-selection Bias **example**

	Disease +	Disease -
Exposure +	300	200
Exposure -	200	300

OR = 2.25

	Disease +	Disease -
Exposure +	240 (80%)	120 (60%)
Exposure -	120 (60%)	180 (60%)

OR = 3.0

Berksons' Bias

- Hospital selective bias
- Patients with two concurrent diseases or health problems are more likely to be admitted to hospital than those with single condition
- Example: people who have peptic ulcer and also who smoke are more likely to be admitted to hospital than people who have either of them.
- A Case-Control study trying to evaluate relationship between smoking and peptic ulcer may find a stronger association between 2.

Incidence – prevalence bias

- Survivorship bias, Neyman's Bias
- Estimate the risk of disease on basis of data collected at a given point in a series of survivors rather than on data gathered during a certain time period in a group of incident cases
- Case-control and cross-sectional study
- Example: case control study to assess the protective effect of physical exercise on MI

Healthy Worker Effect

- Form of selection bias
- General population is often used in occupational studies of mortality, since data is readily available, and they are mostly unexposed
- Example: A comparison between health status of military and civilian population may show a better health status of soldiers because during initial medical examination during which unfit persons are excluded

Bias due to loss to follow up

Differential loss to follow up in a prospective cohort study on oral contraceptives and thromboembolism

Without	TE	Normal
OC+	20	9,980
OC-	10	9,990

RR = 2 (truth)

After 40% loss to follow up

Final	TE	Normal
OC+	8	5,980
OC-	8	5,990

RR = 1 (biased)

Information bias/ measurement bias

- Inadequate means for obtaining information about subjects in the study are inadequate.

TYPES:

1. Non differential mis-classification bias
2. Differential mis-classification bias

Nondifferential mis-classification bias

When errors in exposure or outcome status occur with approximately equal frequency in groups being compared

1. Equally inaccurate memory of exposures in both groups. Example: Case-control study of heart disease and past activity
2. Recording and coding errors in records and databases. Example: ICD9 code used in discharge summaries
3. Using surrogate measures of exposure.
4. Non-specific or broad definitions of exposure or outcome. Example: “do you smoke?” vs (how much, how often, how long) to define exposure to tobacco

example

A Case- Control study comparing CAD cases & controls for history of diabetes.

True relationship

	CAD	Controls
Diabetes	40	10
No	60	90

$$OR = (40 \times 90) / (10 \times 60) = 6$$

With non-differential Misclassification

	CAD	Controls
Diabetes	20	5
No	80	95

$$OR = (20 \times 95) / (5 \times 80) = 4.75$$

Differential mis-classification

When errors in classification of exposure or outcome are more frequent in one group

1. Differences in accurately remembering exposures (unequal). Example: mothers of children with birth defects will remember drugs taken during pregnancy
2. Interviewer or recorder bias. Example: interviewer has subconscious better about hypothesis
3. More accurate information in one of the groups. Example: case-control study with cases from one facility and controls from another with differences in record keeping

Recall Bias

People with disease may remember exposures differently (more or less accurately) than those without disease

To minimize:

1. Use a control group that has a different disease
2. Use questionnaires that are constructed to maximize accuracy and completeness
3. For socially sensitive questions, such as alcohol and drug use, use self-administered questionnaire instead of an interviewer
4. If possible, assess past exposures from pre-existing records

Interviewer bias

Systematic differences in soliciting, recording, or interpreting information

Minimized by

1. Blinding the interviewers if possible
2. Using standardized questionnaires consisting of closed ended, easy to understand questions
3. Training all interviewers to adhere to the question and answer format strictly
4. Obtaining data or verifying data by examining pre-existing records (eg medical records or employment records)

Biases in Case- Control study

- Selection bias
- Information bias
- Bias due to confounding

Biases in Cohort study

- Selection bias
- Follow up bias
- Information bias
- Bias due to confounding
- Post hoc bias

Biases in clinical Trial

- Selection Bias
- Ascertainment bias
- Consent bias
- Dilution bias
- Attrition bias
- Analytical bias
- Publication bias
- Choice of question bias
- Choice of population bias
- Technical bias
- Chance bias

Ascertainment Bias

- This occurs when the person reporting the outcome can be biased.
- Example, of homeopathy study of histamine, showed an effect when researchers were not blind to the allocation but no effect when they were.
- Multiple sclerosis treatment appeared to be effective when clinicians unblinded but ineffective when blinded

Consent bias

- This occurs when consent to take part in the trial occurs AFTER randomisation.
- Most frequent danger in Cluster trials

Dilution bias

- This occurs when the intervention or control group get the opposite treatment. This affects all trials where there is non-adherence to the intervention.
- For example, in a trial of calcium and vitamin D about 4% of the controls are getting the treatment and 35% of the intervention group stop taking their treatment. This will 'dilute' any apparent treatment effect.

Attrition Bias

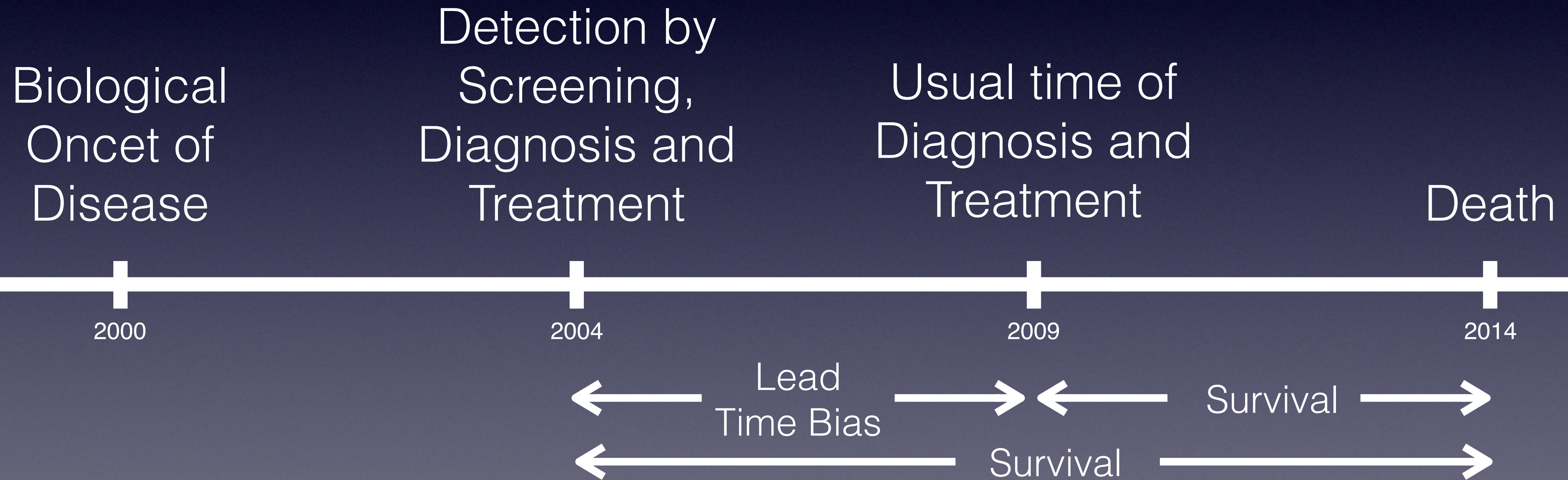
- Usually most trials lose participants after randomisation. This can cause bias, particularly if attrition differs between groups.
- If a treatment has side-effects this may make drop outs higher among the less well participants, which can make a treatment appear to be effective when it is not
- We can avoid some of the problems with attrition bias by using Intention to Treat Analysis, where we keep as many of the patients in the study as possible even if they are no longer 'on treatment'

Biases in screening programmes

- Volunteer bias
- Lead time bias
- Length time bias
- Overdiagnosis bias

Lead time bias

Natural history of disease in hypothetical patient with colon cancer



Length time bias

- Form of selection bias
- Length time bias can occur when lengths of intervals are analysed by selecting intervals that occupy randomly chosen points in time or space
- Example: fast growing tumor has shorter incubation period than slow growing tumor

Overdiagnosis Bias

- Persons who initiate screening program have almost unlimited enthusiasm for the program.
- Even cytologists reading pap smears may become so enthusiastic that they may tend to overread the smears (false positive readings).
- Consequently the abnormal group will be diluted with women who are free of disease.

How to control Selection bias

- Sampling the cases and controls in the same way
- Matching
- Randomization
- Using a population based sample

How to control measurement bias

- Development of explicit, objective criteria for measuring environmental characteristics and health outcomes
- Careful consistent data collection- for example, through use of standardized instruments; objectives, closed ended questionnaires; valid instruments
- Careful consistent use of data instruments- for example, through use of standardized training and instruction manuals, blinding to the extent possible
- Development and application of quality control/ quality assurance procedures
- Use of multiple sources of data
- Data cleaning and coding
- Analysis and adjustment, if necessary, to take account of measurement bias

Confounding

- Mixing or blurring of effects
- Researcher attempts to relate exposure to outcome but actually measures effect of 3rd factor, termed as confounding variable.
- A confounding variable is associated with exposure, affects outcome, but not an intermediate link in chain of causation between exposure and outcome

For a variable to be a confounder

- It should be a known risk factor for the disease or the outcome
- It should be associated with the exposure
- It should not be in direct chain or linked between the exposure and outcome
- It should be differentially distributed in the two group

Control of confounding

- During design of epidemiological study:
 - Randomization
 - Restriction
 - Matching
- During analysis of study:
 - Stratification
 - Statistical modeling or multivariate analysis